

1
2
3
4
5
6
7
8
9
10
11 **IN THE UNITED STATES DISTRICT COURT**
12 **FOR THE NORTHERN DISTRICT OF CALIFORNIA**
13 **OAKLAND DIVISION**

14 JOHN DOE I, et al., on behalf of
15 themselves and all others similarly situated,

16 Plaintiffs,

17 v.

18 GOOGLE LLC,

19 Defendant.
20
21
22
23
24
25
26
27
28

Case No. 5:23-cv-02431-BLF

**DECLARATION OF DR. TIMOTHY
LIBERT IN SUPPORT OF PLAINTIFFS'
MOTION FOR PRELIMINARY
INJUNCTION**

DECLARATION OF DR. TIMOTHY LIBERT

I, DR. TIMOTHY LIBERT, hereby declare under penalty of perjury:

1. I am the owner of webXray LLC, a consulting firm that works on issues involving the presence, and functioning of, code on websites which generate network requests that may be used to track and profile users.

2. I submit this declaration in support of Plaintiffs' Motion for Preliminary Injunction.

3. I have personal knowledge of the facts set forth herein and, if called as a witness, could and would testify competently to them.

4. I reserve the right to modify, supplement or otherwise amend my statements, analyses, and conclusions as new and additional information becomes available to me.

I. QUALIFICATIONS

5. I am highly experienced in the area of web technologies and privacy. I have been designing, coding, and analyzing websites for the past 24 years, and I am fluent in several programming languages and data analysis techniques. I received my Ph.D. from the University of Pennsylvania in 2015 where I pursued interdisciplinary coursework and research in communications policy, computer science, law, business, and medicine. I completed my post-doctoral training at the University of Oxford where I studied the impact of Europe's General Data Protection Regulation on news websites. I was a Special Faculty Instructor at Carnegie Mellon University for three years where I taught several courses in the world's only Privacy Engineering program and conducted research into web tracking at the CyLab Security and Privacy Institute.

6. Currently, I am the owner of webXray LLC, which is a data protection consultancy focused on measuring and documenting privacy violations on the web.

7. Attached as **Exhibit A** is a current copy of my curriculum vitae.

II. ASSIGNMENT AND SUMMARY CONCLUSIONS

8. I have been asked to determine whether it is possible to identify health care websites (defined below) where the inclusion of Google source code is causing the interception and transmission of patient information to Google, and, if so, to perform an analysis of these websites and the corresponding prevalence of Google tracking and interception of patient information. None

1 of my analysis involved the collection, use, or examination of any information which is not publicly
2 available.

3 9. As set forth below, I was able to compile a subset of health care websites and then
4 conduct testing to identify the websites where the inclusion of Google source code caused the
5 interception and transmission of patient information to Google. My results showed that Google
6 source code was present on 87% of the health care websites, with Google Analytics being present
7 67% of the time, Google Ads 58% of the time, Google Display Ads 59% of the time.

8 10. This comprehensive and indiscriminate data collection means those seeking
9 information related to the mental health of their children, medical help after a sexual assault, and
10 loans to pay medical bills have these deeply sensitive topics – and many more – transmitted to
11 Google. For example, and as described in detail below, my results showed Google presence on the
12 following web pages:

- 13 • <https://www.invega.com/adolescent-childhood-schizophrenia.html>
- 14 • <https://www.creekhealth.org/services/sexual-assault-nurse-exams/>
- 15 • [https://www.effinghamhealth.org/billing-insurance/medical-credit-card-loan-](https://www.effinghamhealth.org/billing-insurance/medical-credit-card-loan-options/)
16 [options/](https://www.effinghamhealth.org/billing-insurance/medical-credit-card-loan-options/)

17 The process of conducting this analysis was straightforward and uses techniques which are well
18 known to qualified professionals.

19 11. Having determined that it is relatively easy to identify a web property with Google
20 source code, I next engaged in a more in-depth analysis, looking at every page within a web
21 property (i.e. the “site map”) to determine the prevalence of Google source code inclusion within a
22 particular web property. For this analysis, I examined the following three sample hospitals – Kaiser
23 Permanente, Gunderson Health Care (“Gundersen”), and MD Mercy Health Services (“MD
24 Mercy”). It is my understanding that each of these hospitals corresponds with some of the named
25 plaintiffs in this case.

26 12. As set forth below, my analysis reveals that the inclusion of Google source code is
27 consistently present on pages that contain private and sensitive information. Notably, the URLs
28 for these pages – which are conveyed to Google – provide clear descriptions of what a patient is

1 looking for, as seen in the following examples:

- 2 • [https://about.kaiserpermanente.org/health-and-wellness/our-care/seeking-early-](https://about.kaiserpermanente.org/health-and-wellness/our-care/seeking-early-detection-for-pancreatic-cancer)
- 3 [detection-for-pancreatic-cancer](https://about.kaiserpermanente.org/health-and-wellness/our-care/seeking-early-detection-for-pancreatic-cancer)
- 4 • <https://www.gundersenhealth.org/services/mychart-e-visit>
- 5 • <https://mdmercy.com/find-a-doctor/kelly-alexander-md/request-appointment>

6 The above URL is merely an example of a vast universe of such URLs which contain clear
7 descriptions of the content of the patient's inquiry.

8 13. Further, because the inclusion of Google source code is pervasive within the web
9 property (i.e. presence on almost every single web page), the sequence of network transmissions
10 being sent to Google would effectively disclose nearly every step, navigation and communication
11 that is taken by a particular patient.

12 **III. IT IS POSSIBLE TO IDENTIFY THE HEALTH CARE WEBSITES WHICH ARE** 13 **SUBJECTED TO GOOGLE SOURCE CODE**

14 **A. METHODOLOGY**

15 14. In this section I discuss how automated web privacy analysis can be used to
16 investigate and identify health care web-properties that contain Google source code. Specifically,
17 as to the inclusion of Google source code, I analyzed for the following: Google Analytics, Google
18 Ads, Google Display Ads, YouTube, Google Tag Manager, and Google APIs.

19 15. The first step in my analysis was to select a set of properties to investigate. To that
20 end, I attempted to compile as many web-properties as possible from the following categories:
21 hospitals, pharmacies, prescription drugs, and health insurers. To be clear, the web-properties that
22 I was able to compile is not intended to represent the full list of "Health Care Providers," as that
23 phrase is defined in Plaintiffs' complaint. *See* Dkt. 1, fn. 1. Rather, this was an exercise to see how
24 feasible it is to compile the list and then analyze the presence of Google source code.

25 16. With respect to hospitals, I was given a list of 3,362 hospital web-properties from
26 Plaintiffs' counsel.

27 17. With respect to pharmacies, I researched and compiled a list of 101 pharmacies.
28

1 18. With respect to prescription drug sites, I looked to the U.S. Food and Drug
2 Administration list of approved medications as a starting point and then found the homepages for
3 each drug, ending up with 1,765 prescription drug web properties.

4 19. With respect to health insurers, I used a variety of web search methods to find web
5 properties which offered coverage to individuals for medical, dental, or vision, thereby producing
6 786 sites.

7 20. The second step in my analysis was to detect tracking code, cookies, and network
8 traffic which expose user data. I have been analyzing privacy issues on web-properties for over a
9 decade, using a tried-and-true methodology that I have perfected over many years. At a high level,
10 I use software which instructs the web browser to open a specific URL or set of URLs (e.g. web
11 pages within a specific web property). When a page is open, basic user actions such as scrolling
12 and clicking to additional pages on the same site are performed. At the conclusion of the page
13 interaction, the software processes, records, and stores all source code, cookies, network traffic,
14 and other pertinent information in a searchable database. I am then able to easily generate summary
15 reports of the data as well as identify specific lines of code, cookies, and the like.¹ For example, I
16 can easily review the summary reports to identify network traffic which downloads Google source
17 code to a user's browser, thereby exposing their browsing behaviors to Google.

18 **B. ANALYSIS**

19 21. Proceeding under the above methodology, I ran the list of web properties through
20 the tool, instructing the program to analyze the home page and a random set of web pages from
21 within each health care web property. I then reviewed the summary report for the presence of
22 network transmissions to Google that exposed user information, identifiers, and browsing activity
23 to Google.

24 ¹ My webXray technology has been in development since 2015 and the corresponding methodology
25 has been peer reviewed in the most competitive academic venues in the world and I have numerous
26 publications which provide evidence the approach is sound. Some notable examples include Libert,
27 Timothy. "Privacy implications of health information seeking on the web." *Communications of the*
28 *ACM* 58.3 (2015): 68-77.; Libert, Timothy "An automated approach to auditing disclosure of third-
party data collection in website privacy policies." *Proceedings of the 2018 World Wide Web*
Conference. 2018.; and McCoy, Matthew S., et al. "Prevalence of third-party tracking on COVID-
19-related web pages." *Jama* 324.14 (2020): 1462-1464.

1 22. Specifically, I investigated network transmissions to the following domains that
2 belong to the following Google products:

3 a. Google Display Ads, also known as DoubleClick, which is associated with the
4 following domains:

- 5 i. 2mdn.net
- 6 ii. doubleclick.com
- 7 iii. doubleclick.net
- 8 iv. doubleclickbygoogle.com

9 b. Google Ads, which is associated with the following domains:

- 10 i. google.com
- 11 ii. googleadservices.com
- 12 iii. googlesyndication.com

13 c. Google Analytics, which is associated with the following domains:

- 14 i. google-analytics.com
- 15 ii. analytics.google.com

16 d. Google APIs, which is associated with googleapis.com domain.

17 e. Google Tag Manager, which is associated with the following domains:

- 18 i. googletagmanager.com
- 19 ii. googletagservices.com

20 f. YouTube, which is associated with the following domains:

- 21 i. youtube.com
- 22 ii. youtubetv.com
- 23 iii.youtu.be
- 24 iv. youtube-nocookie.com
- 25 v. ytimg.com

26 23. Based on my review of the summary reports generated, and leveraging the above
27 list of Google domains, I am able to conclude as follows:

28 a. 87% of all the web properties analyzed made requests to download code that

1 generated network traffic which exposed user information, identifiers, and
2 browsing activity to Google.

3 b. 67% of all web properties analyzed made requests to download code that
4 generated network traffic which exposed user information, identifiers, and
5 browsing activity to Google Analytics.

6 c. 58% of all web properties analyzed made requests to download code that
7 generated network traffic which exposed user information, identifiers, and
8 browsing activity to Google Ads.

9 d. 59% of all web properties analyzed made requests to download code that
10 generated network traffic which exposed user information, identifiers, and
11 browsing activity to Google Display Ads.

12 e. 69% of all web properties analyzed made requests to download code that
13 generated network traffic which exposed user information, identifiers, and
14 browsing activity to Google Tag Manager.

15 f. 19% of all web properties analyzed made requests to download code that
16 generated network traffic which exposed user information, identifiers, and
17 browsing activity to YouTube.

18 g. 66% of all web properties analyzed made requests to download code that
19 generated network traffic which exposed user information, identifiers, and
20 browsing activity to Google APIs.

21 **IV. ANALYSIS OF GOOGLE SOURCE CODE ON SPECIFIC WEB PAGES WITHIN**
22 **SAMPLE WEB PROPERTIES**

23 24. Whereas my analysis above relied on loading the home page and a random selection
24 of internal pages, to conduct an even more in-depth review of the inclusion of Google source code
25 on a web property, I analyzed a large volume of pages on three sample hospital web properties:
26 Kaiser Permanente, Gundersen, and MD Mercy.

27 25. In the cases of Gundersen and MD Mercy I used their publicly available
28 “sitemap.xml” file which provides a list of URLs that are available on these respective web

properties. A sitemap “is a file where you provide information about the pages, videos, and other files on your site, and the relationships between them.”² A publicly available site map does not exist for Kaiser Permanente, thus I iteratively extracted links from a given page and loaded those pages to find more links.

26. While a random sampling of pages provides insights into the presence or absence of Google on a specific site, a deeper analysis of internal pages may reveal the specific URLs, and patient interests, sent to Google.

27. For example, on the Kaiser web property, network transmissions are made to Google on the following web page URLs that contain clear descriptions of the content of the patient’s communications:

- <https://about.kaiserpermanente.org/health-and-wellness/our-care/remote-patient-monitoring-improves-care-for-chronic-disease>
- <https://about.kaiserpermanente.org/health-and-wellness/our-care/seeking-early-detection-for-pancreatic-cancer>
- <https://about.kaiserpermanente.org/health-and-wellness/our-care/cardiac-care>
- <https://well.kaiserpermanente.org/physicians/raman-luhadiya/>

28. On the Gundersen web property, network transmissions are made to Google on the following web page URLs that contain clear descriptions of the content of the patient’s communications:

- <https://www.gundersenhealth.org/our-system/request-an-appointment>
- <https://www.gundersenhealth.org/pay-my-bill/online-bill-pay-help>
- <https://www.gundersenhealth.org/services/mychart-e-visit>
- <https://www.gundersenhealth.org/services/non-surgical-weight-loss/non-surgical-weight-loss-seminar/weight-management-obesity>

29. On the MD Mercy web property, network transmissions are made to Google on the following web page URLs that contain clear descriptions of the content of the patient’s communications:

² “Google Search Central: Learn about sitemaps”

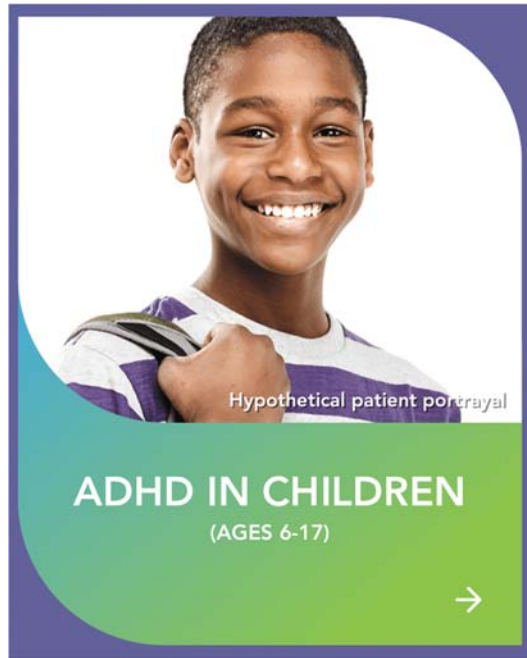
https://developers.google.com/search/docs/crawling-indexing/sitemaps/overview?hl=en&visit_id=638217792861284157-117525059&rd=1

- 1 • <https://mdmercy.com/find-a-doctor/kelly-alexander-md/request-appointment>
- 2 • <https://mdmercy.com/mercy-services/cancer-institute/medical-oncology-and-hematology/appointments-and-contact/request-appointment>
- 3 • <https://mdmercy.com/mercy-services/cancer-institute/melanoma-and-skin-cancer-center>
- 4 • <https://mdmercy.com/mercy-services/cancer-institute/surgical-oncology/diagnostic-and-specialty-services?listPage=1>

5
6 30. Analysis of the site maps, and Google’s prevalence throughout a particular web
7 property, also demonstrates that Google’s access to individuals’ engagement with their Health Care
8 Provider web properties – their inquiries, communications, actions, and navigation history – is
9 incredibly pervasive. The analysis of Kaiser Permanente, Gundersen and MD Mercy site maps
10 reveals that Google is present on nearly all web pages. The net effect is that Google does not get
11 just a snapshot of an individuals’ interaction and communications, but rather an almost unrestricted
12 view and access to the full navigation and browsing history undertaken by individuals with their
13 health care providers.

14 31. Further, my analysis provides evidence of transmission to Google of “click-through”
15 behavior, i.e., a transmission that communicates that an individual is in fact transitioning from one
16 page to another. My software simulates user activity by clicking links and scrolling on a page, in
17 tandem, it monitors how such activities may be monitored by outside parties in real time.

18 32. For example, when a user on the home page of the ADHD medication Vyvanse
19 clicks on the link for “ADHD IN CHILDREN” (shown below), the text “Click” and “ADHD IN
20 CHILDREN” is sent to Google Analytics.
21
22
23
24
25
26
27
28



33. In fact, as demonstrated, this “click through” action is captured simultaneously by Google Analytics and Google Tag Manager. The specific transmissions to each of these Google products are as follows:

Google Analytics:

[https://www.google-analytics.com/collect?v=1&_v=j100&a=1329054035&t=event&ni=0&_s=1&cd=https%3A%2F%2Fwww.vyvansepro.com%2F&dl=https%3A%2F%2Fwww.vyvansepro.com%2F&dp=%2F&ul=en-us&de=UTF-8&dt=Vyvanse%C2%AE%20Uses%2C%20Dosage%2C%20Clinical%20Data%2C%20%26%20Resources%3A%20For%20HCPs&sd=24-bit&sr=1280x720&vp=&jc=0&ec=Clicks&ea=Click&el=Clicks%3AClick%3AADHD%20IN%20CHILDREN%20\(AGES%206-17\)%20Hypothetical%20patient%20portrayal%3Ahttps%3A%2F%2Fwww.vyvansepro.com%2Fadhd%2Fadhd-in-children&_u=aADAAEABAAAAACAAI~&jid=&gjid=&cid=1342691313.1686082273&tid=UA-2357337-11&_gid=2091258812.1686082273>m=45He3650n81NF4GCVP&z=1940072035](https://www.google-analytics.com/collect?v=1&_v=j100&a=1329054035&t=event&ni=0&_s=1&cd=https%3A%2F%2Fwww.vyvansepro.com%2F&dl=https%3A%2F%2Fwww.vyvansepro.com%2F&dp=%2F&ul=en-us&de=UTF-8&dt=Vyvanse%C2%AE%20Uses%2C%20Dosage%2C%20Clinical%20Data%2C%20%26%20Resources%3A%20For%20HCPs&sd=24-bit&sr=1280x720&vp=&jc=0&ec=Clicks&ea=Click&el=Clicks%3AClick%3AADHD%20IN%20CHILDREN%20(AGES%206-17)%20Hypothetical%20patient%20portrayal%3Ahttps%3A%2F%2Fwww.vyvansepro.com%2Fadhd%2Fadhd-in-children&_u=aADAAEABAAAAACAAI~&jid=&gjid=&cid=1342691313.1686082273&tid=UA-2357337-11&_gid=2091258812.1686082273>m=45He3650n81NF4GCVP&z=1940072035)

Google Tag Manager

https://www.googletagmanager.com/a?id=GTM-NF4GCVP&v=3&t=t&pid=689399374&cv=12&rv=3650&tc=143&es=1&e=gtm.click&e_id=45&u=AgAAAAAAAAAAAAACCA&h=Ag&pcr=41.41.41.41&z=0

34. These transmissions once again demonstrate the extent of Google’s access to patients’ health information and communications with their health care providers.

1 35. I declare under penalty of perjury under the laws of the United States of America
2 that the foregoing is true and correct.

3 Executed this 12th day of June 2023 at Sunnyvale, California.

4 

5
6 /s/ _____
7 Dr. Timothy Libert